Report of Indicraft Vintage pvt.Ltd

As per the analysis of the data this report will help to visualise ,and analysis data , and quickly see progress against key performance indicator(KPIs) and other matrices.

The dataset contains various features related to properties, including identifiers, building and zoning information, lot details, road and lane types, property shape, utility types, sale details, and prices.

The dataset contains 81 columns and 1459 rows. Key columns include various attributes related to the properties and their sales prices. Some columns have missing values. Here's a summary of the dataset's structure:

- **Id**: Unique identifier for each property.
- **Building_Class**: Type of building.
- **Zoning_Class**: Classification of the zone.
- **Lot_Extent**: Lot size in square feet.
- **Lot_Size**: Area of the lot.
- **Road_Type**: Type of road access.
- **Lane_Type**: Type of lane access (many missing values).
- **Property_Shape**: Shape of the property.
- **Land_Outline**: Flatness of the property.
- **Utility_Type**: Type of utilities available.
- **House_Type**: Type of house.
- **House_Design**: Design of the house.
- **Overall_Material**: Overall material and finish quality.
- **House_Condition**: Condition of the house.
- **Construction_Year**: Year the house was built.
- **Remodel_Year**: Year the house was remodeled.
- **Roof_Design**: Type of roof design.
- **Exterior_Material**: Type of exterior material.
- **Foundation_Type**: Type of foundation.
- **Basement_Height**: Height of the basement.
- **Sale_Price**: Sale price of the property.

Let's proceed with exploratory data analysis (EDA) to uncover trends, patterns, and insights. The key steps will include:

1. **Summary Statistics**: Understanding the central tendencies and dispersion of numerical features.
2. **Missing Values**: Identifying and handling missing values.

3. **Distribution of Sale Prices**: Visualising the distribution of property sale prices.
4. **Correlations**: Analysing the correlation between numerical features and sale prices.
5. **Categorical Features**: Examining the impact of categorical features on sale prices.
6. **Visualisations**: Creating visualisations to present insights effectively.

Summary Statistics

Here are the summary statistics for some key numerical features:

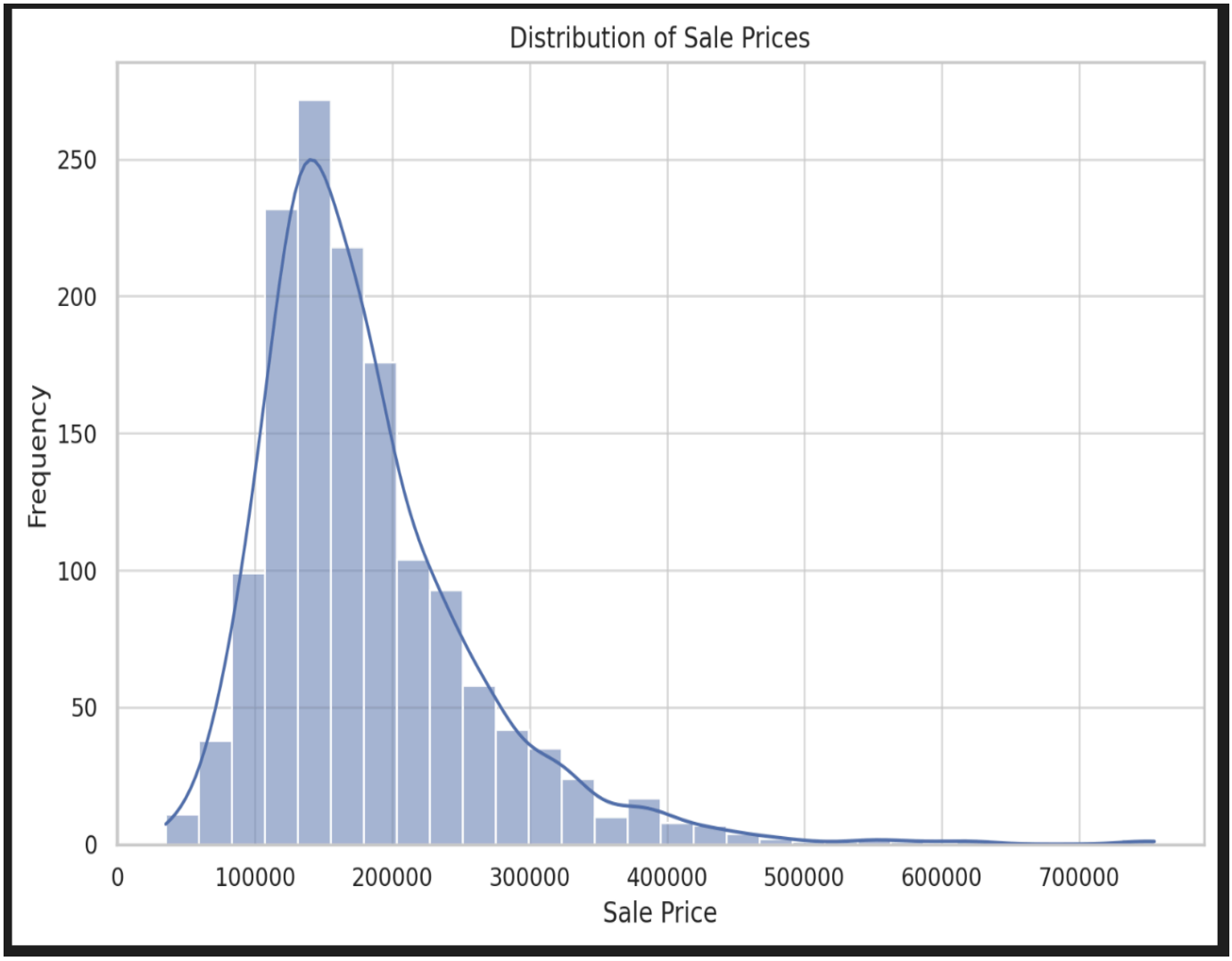
- **Lot_Extent**: Mean = 70.05, Std = 24.29, Min = 21.00, Max = 313.00
- **Lot_Size**: Mean = 10,517.23, Std = 9,984.68, Min = 1,300.00, Max = 215,245.00
- **Overall_Material**: Mean = 6.10, Std = 1.38, Min = 1.00, Max = 10.00
- **House_Condition**: Mean = 5.58, Std = 1.11, Min = 1.00, Max = 9.00
- **Sale_Price**: Mean = 180,944.10, Std = 79,464.92, Min = 34,900.00, Max = 755,000.00

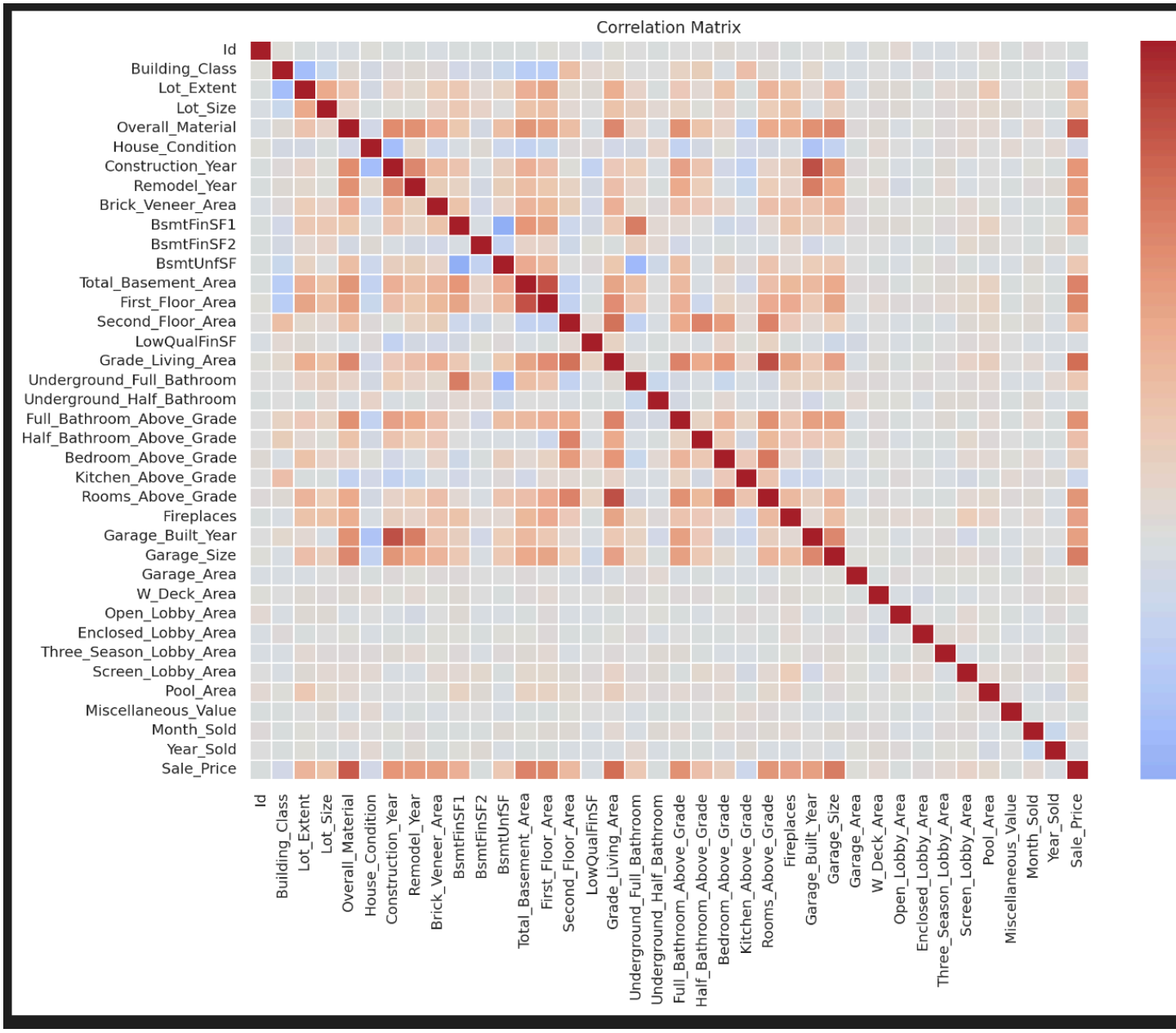
Missing Values

Some features have missing values:

- **Lot_Extent**: 259 missing values
- **Lane_Type**: 1,368 missing values
- **Pool_Quality**: 1,450 missing values
- **Fence_Quality**: 1,173 missing values

Visualisation of the distribution of sale prices and explore correlations between numerical features and sale prices.


Distribution of Sale Prices
Frequency
Sale Price
0
50
100
150
200
250
0
100000
200000
300000
400000
500000
600000
700000

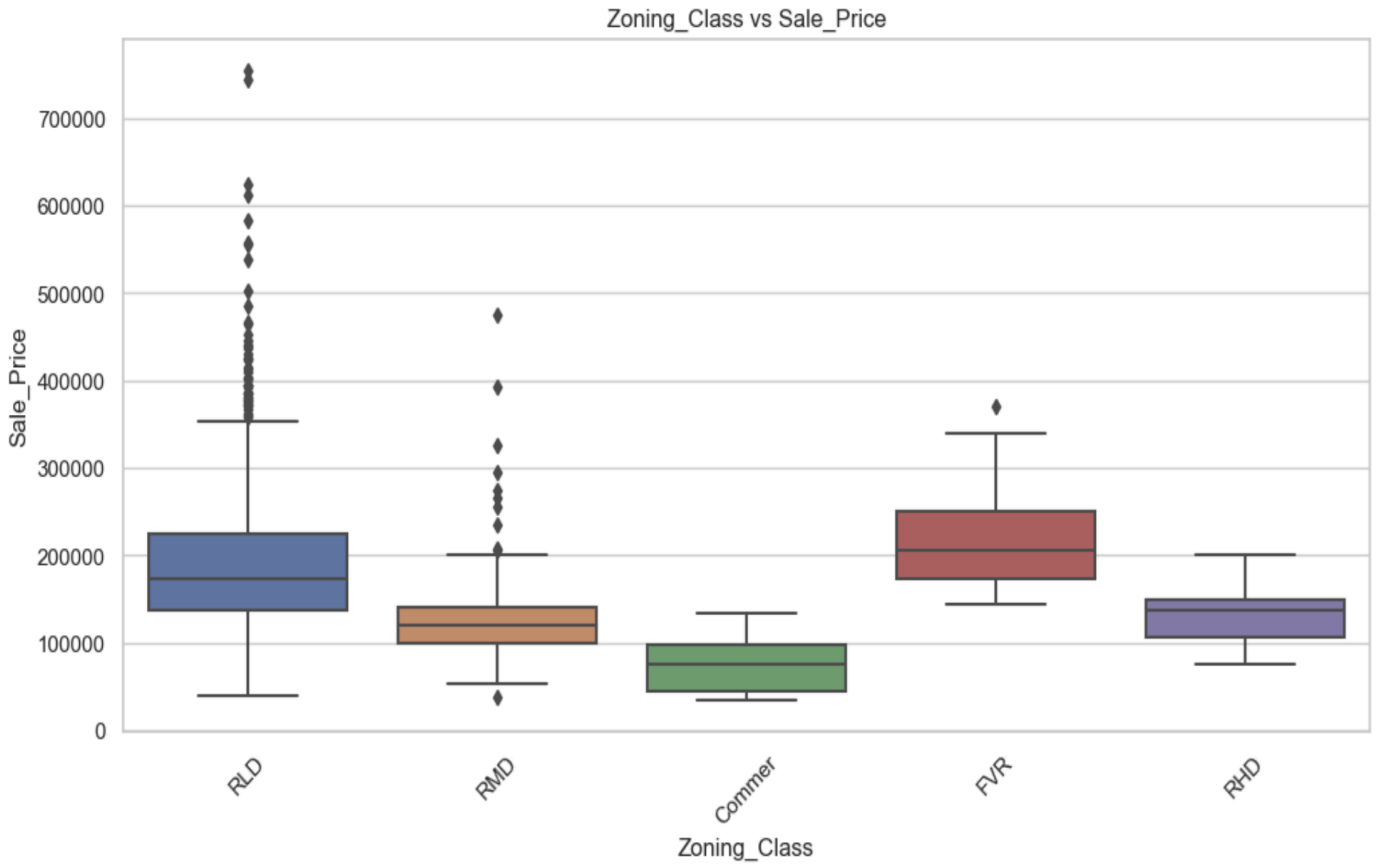


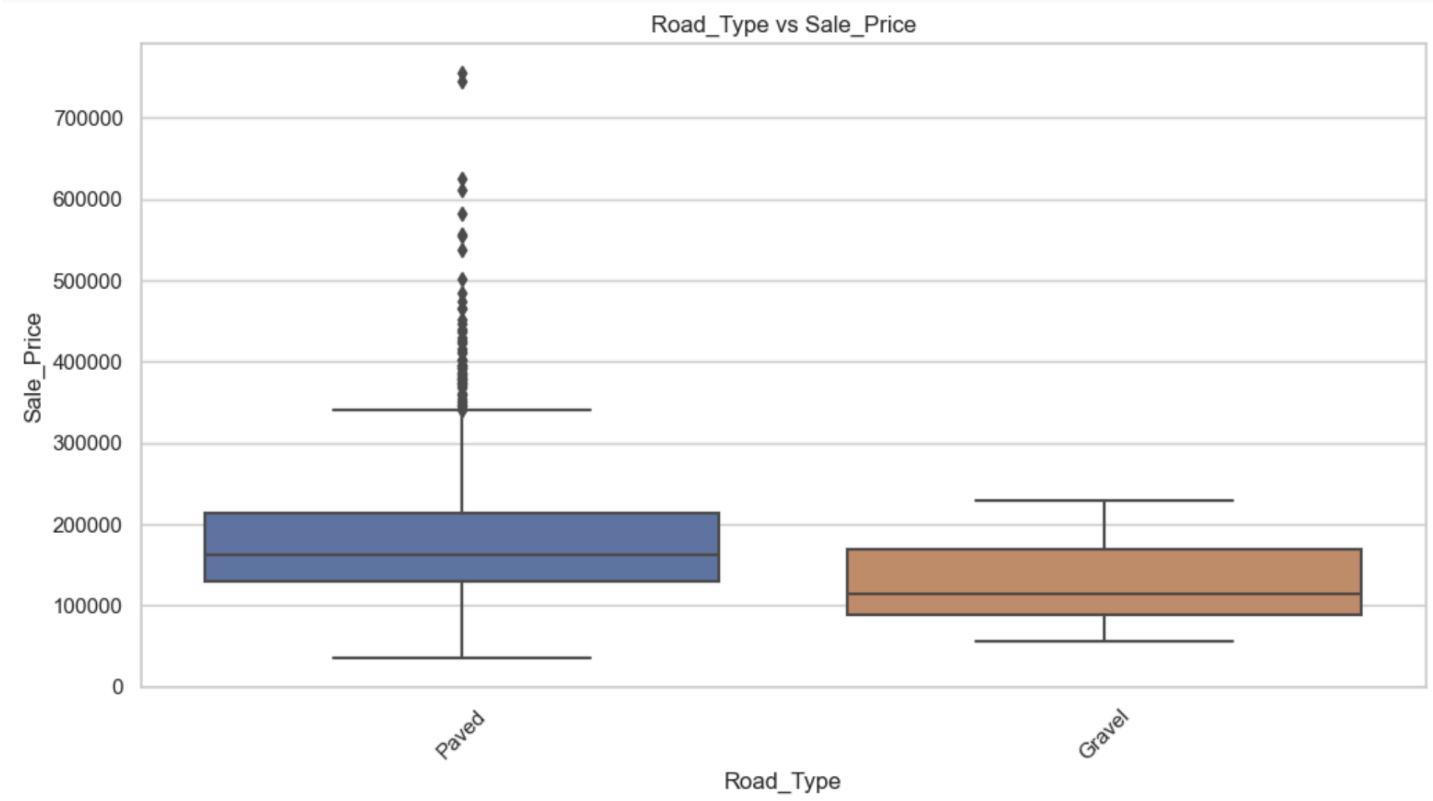
Insights from the EDA

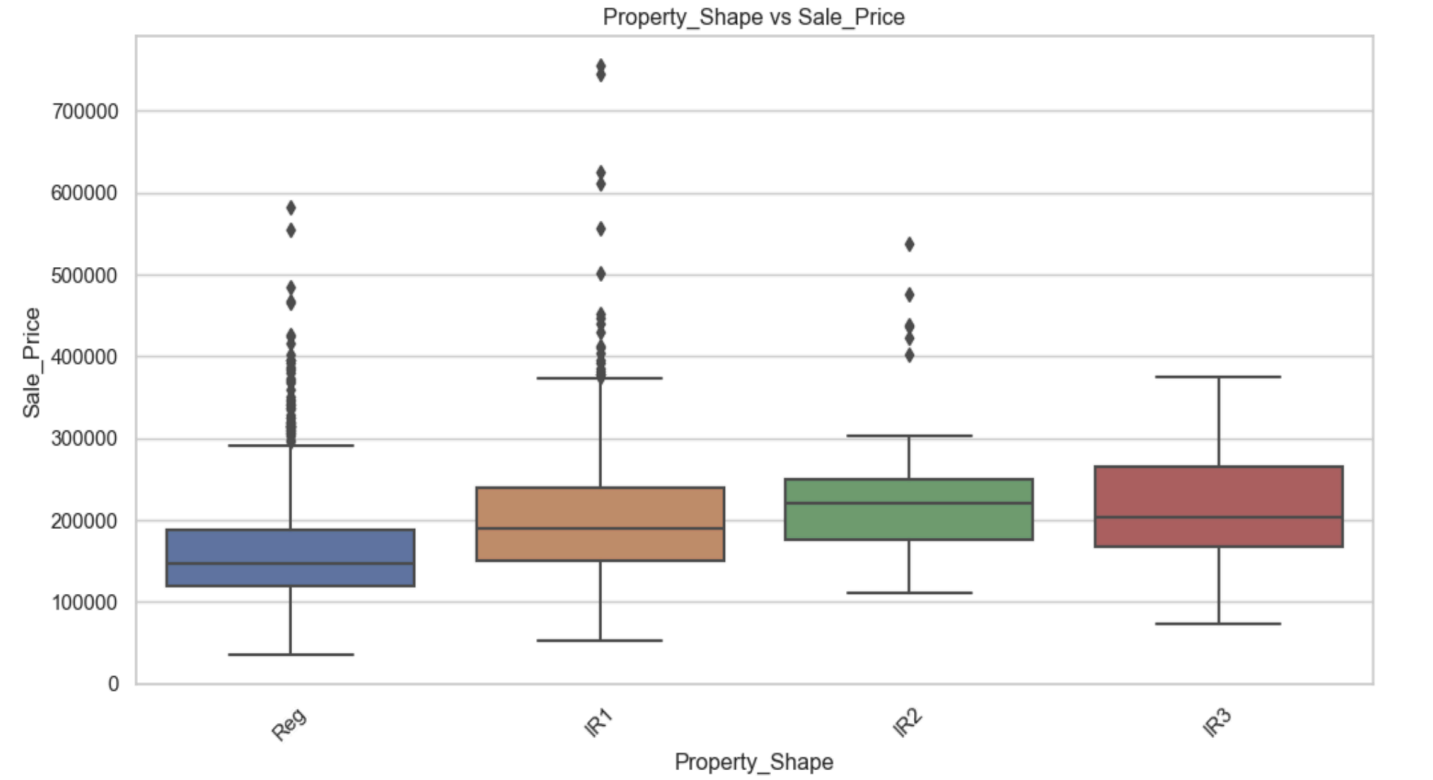
1. **Distribution of Sale Prices**:
    - The sale prices are right-skewed, with most properties selling for prices below $200,000.
2. **Correlations with Sale Prices**:
    - The following features have the highest positive correlations with sale prices:
        - **Overall Material**: 0.79
        - **Grade Living Area**: 0.71
        - **Garage Size**: 0.64

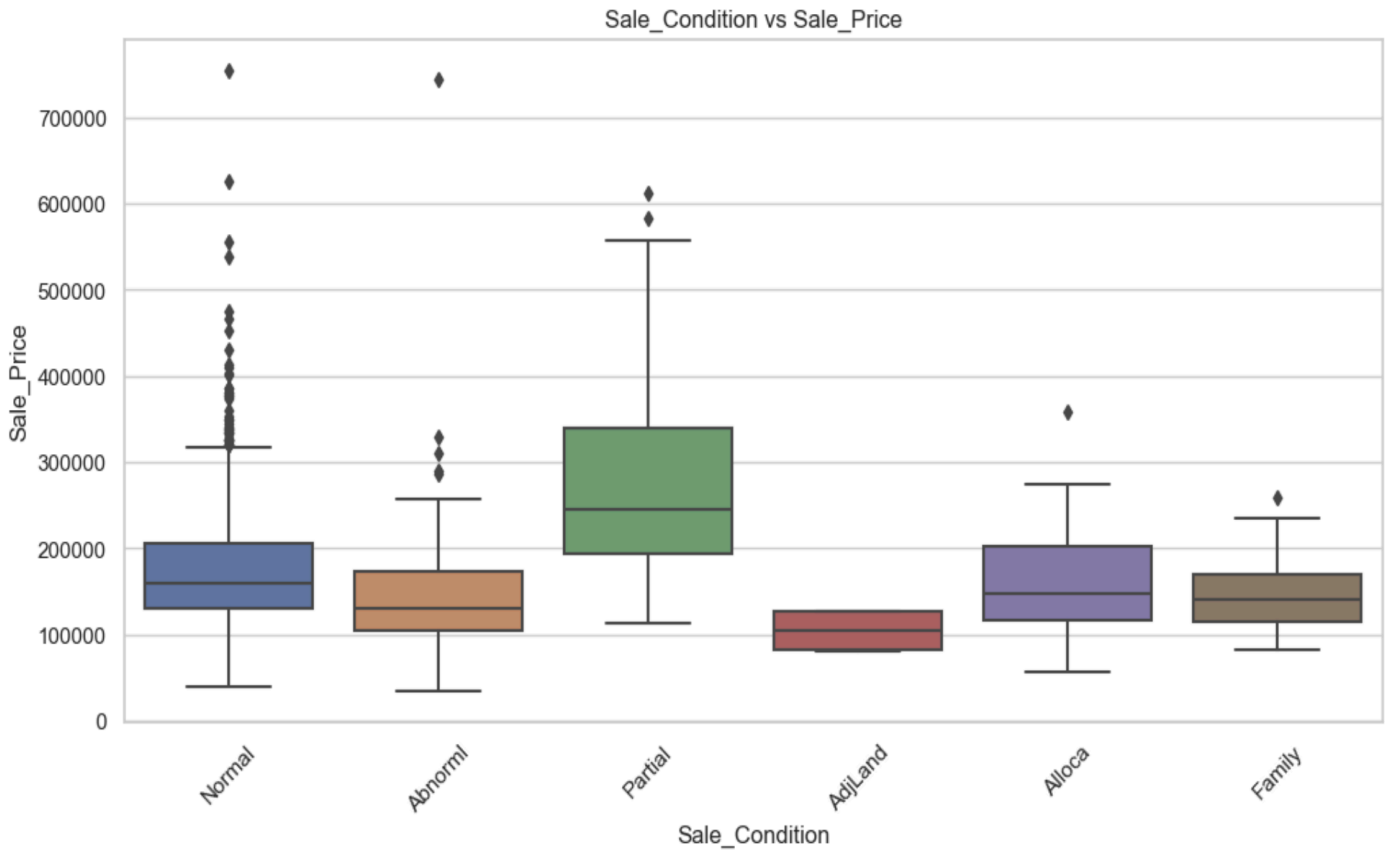
        - **Total Basement Area**: 0.61
        - **First Floor Area**: 0.61
    - Negative correlations:
        - **Kitchen Above Grade**: -0.14
        - **Building Class**: -0.08
        - **House Condition**: -0.08
3. **Visualisations**:
    - The heatmap of the correlation matrix provides an overview of relationships among numerical features. The strongest correlations with sale prices are evident in features related to the size and quality of the property.

Analysis of the impact of some categorical features on sale prices through visualisations. We'll focus on features like (`Zoning_Class`, `Road_Type`, `Property_Shape`, and `Sale_Condition`)




Road_Type vs Sale_Price
Sale_Price
700000
600000
500000
400000
300000
200000
100000
0
Paved
Gravel
Road_Type

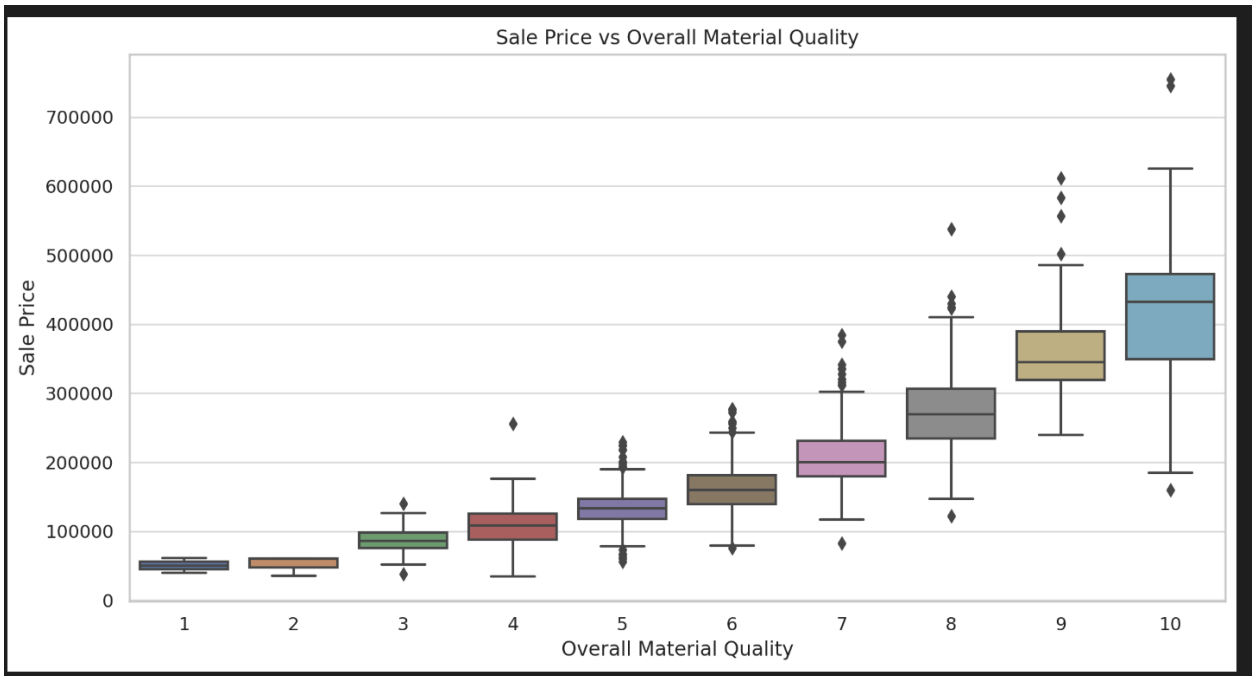

Property_Shape vs Sale_Price
Sale_Price
700000
600000
500000
400000
300000
200000
100000
0
Reg
IR1
IR2
IR3
Property_Shape

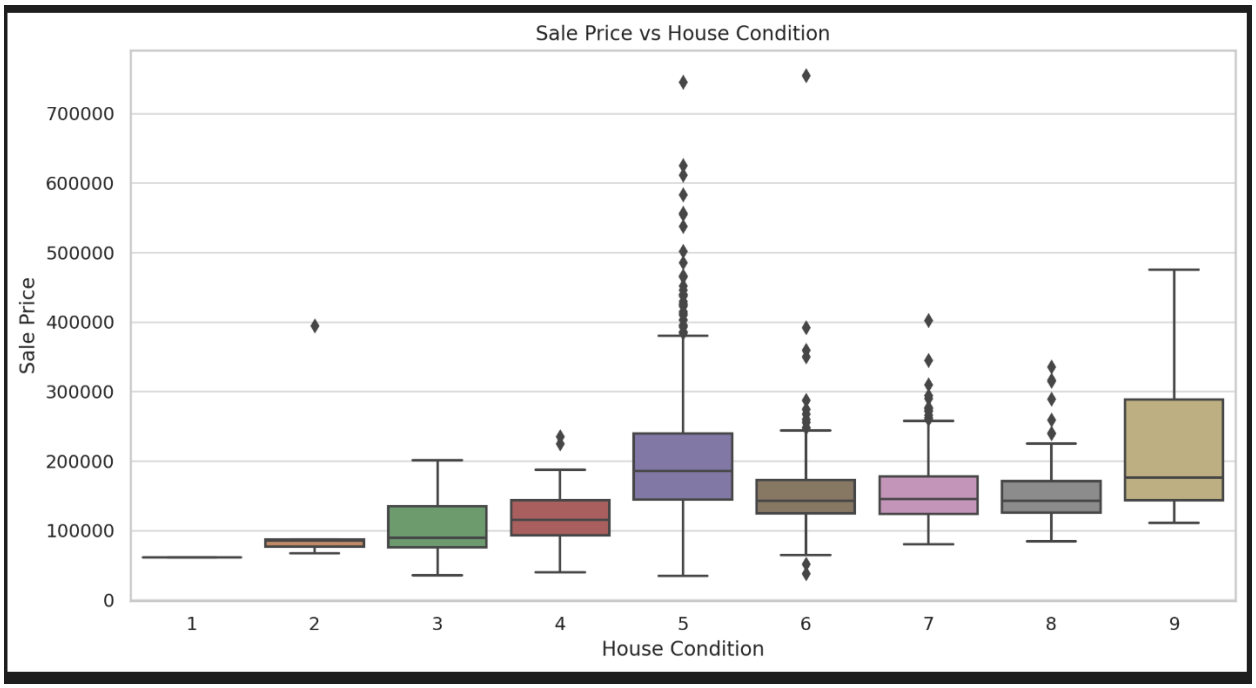


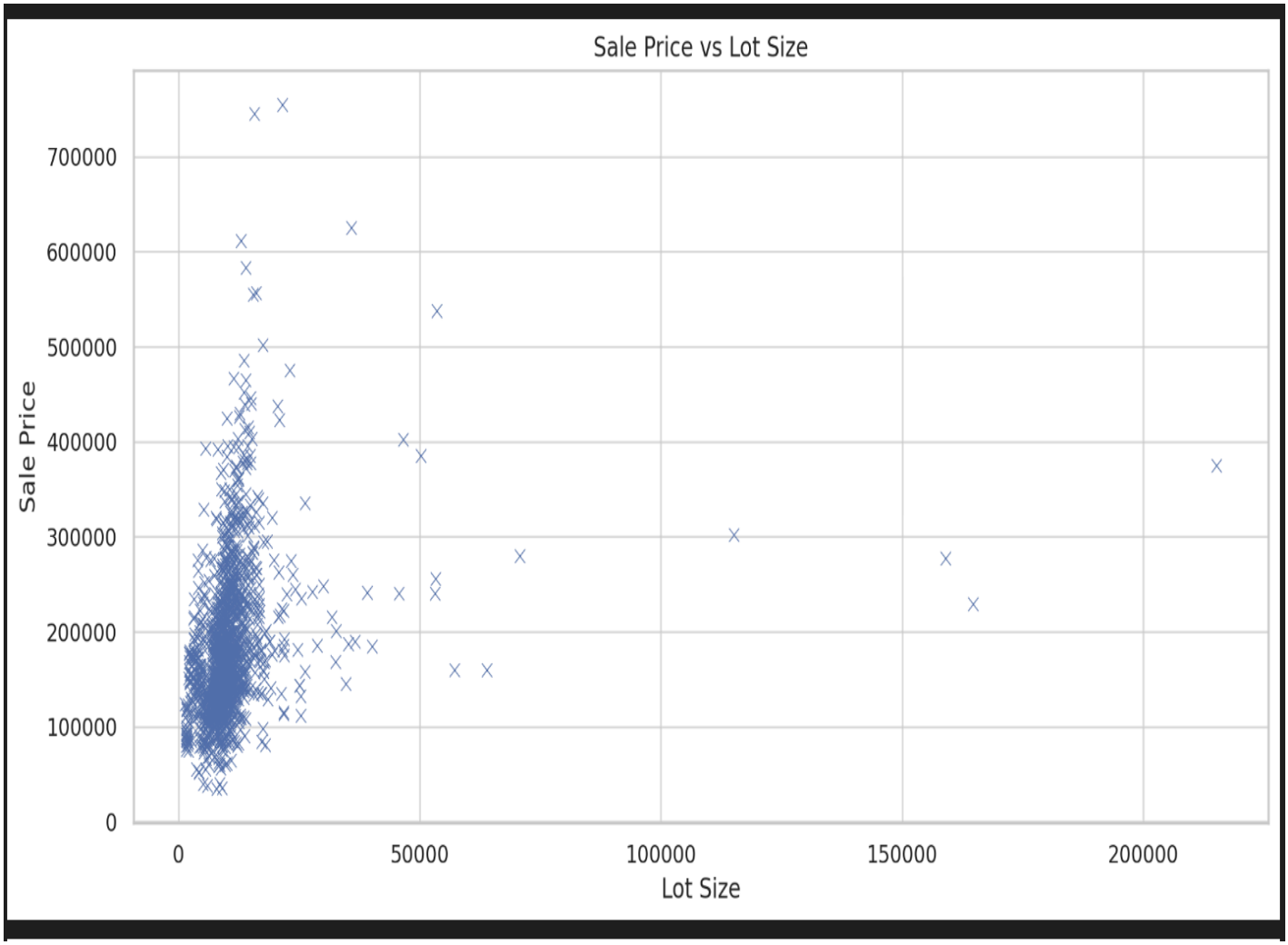
Above are the detailed analysis,of how categorical features like `Zoning_Class`, `Road_Type`, `Property_Shape`, and `Sale_Condition` impact sale prices using boxplots or other visualisations.

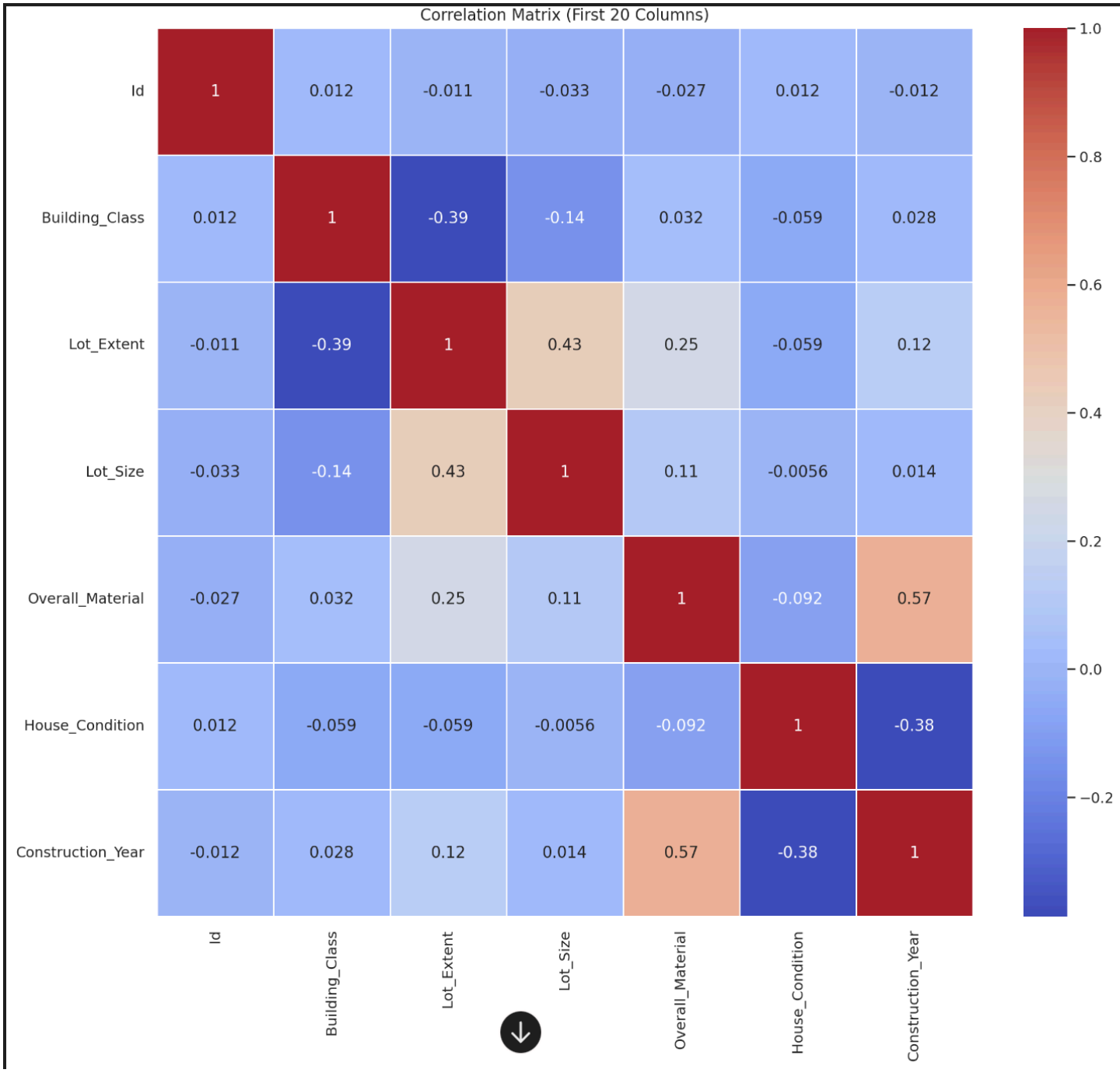
Summarising findings and insights.

Some visualisations to identify key trends and insights, focusing on the relationship between the sale price and other variables.






Sale Price vs Lot Size
Sale Price
700000
600000
500000
400000
300000
200000
100000
0
0
50000
100000
150000
200000
Lot Size


Correlation Matrix (First 20 Columns)
Id
Building_Class
Lot_Extent
Lot_Size
Overall_Material
House_Condition
Construction_Year
1 0.012 -0.011 -0.033 -0.027 0.012 -0.012
0.012 1 -0.39 -0.14 0.032 -0.059 0.028
-0.011 -0.39 1 0.43 0.25 -0.059 0.12
-0.033 -0.14 0.43 1 0.11 -0.0056 0.014
-0.027 0.032 0.25 0.11 1 -0.092 0.57
0.012 -0.059 -0.059 -0.0056 -0.092 1 -0.38
-0.012 0.028 0.12 0.014 0.57 -0.38 1
1.0
0.8
0.6
0.4
0.2
0.0
−0.2

Report: Analysis of Property Prices

Summary of Key Findings and Insights

1. **Overall Material Quality vs. Sale Price**
    - Higher overall material quality generally correlates with higher sale prices.
    - There is a noticeable increase in median sale price with better material quality.
2. **House Condition vs. Sale Price**
    - Sale prices tend to be higher for houses in better condition.
    - The median sale price shows an upward trend as the house condition improves.
3. **Distribution of Sale Price**
    - Sale prices are right-skewed with most properties priced between $100,000 and $300,000.
    - There are some high-value outliers beyond $500,000.
4. **Sale Price vs. Lot Size**
    - There is a positive correlation between lot size and sale price, but with significant variance.
    - Larger lots tend to have higher sale prices, though the relationship is not strictly linear.
5. **Correlation Analysis (First 20 Columns)**
    - The correlation heatmap reveals several relationships among the first 20 variables.
    - Strong positive correlations exist between variables such as `Overall_Material` and `Sale_Price`, `House_Condition` and `Sale_Price`.

Visualisations

1. **Sale Price vs. Overall Material Quality**
2. **Sale Price vs. House Condition**
3. **Distribution of Sale Price**
4. **Sale Price vs. Lot Size**
5. **Correlation Matrix (First 20 Columns)**

Recommendations

- **For Buyers: Consider properties with higher overall material quality and better condition, as these tend to offer better value for money.**
- **For Sellers: Investing in improving the material quality and condition of your property can significantly enhance its market value.**
- **For Investors: Larger lots offer higher potential returns but with more variance. Focus on properties with high-quality materials and good conditions to minimise risk.**

This analysis provides a foundational understanding of the key factors affecting property prices, which can guide decisions for buyers, sellers, and investors.